PyTorch
CONFERENCE 2025

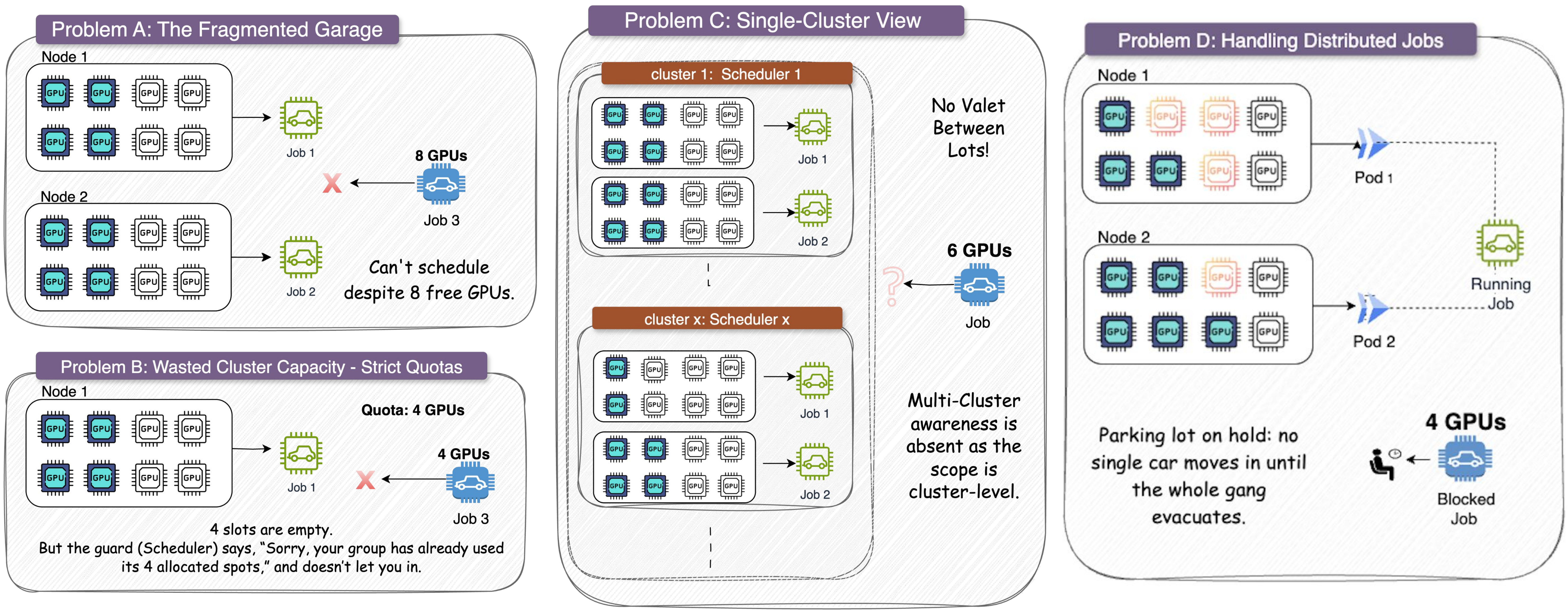
Scheduling ML Jobs at Scale

Manvi Gupta
(Computer Scientist 1, Adobe Systems)

Problem Space

The GPU Parking Problems

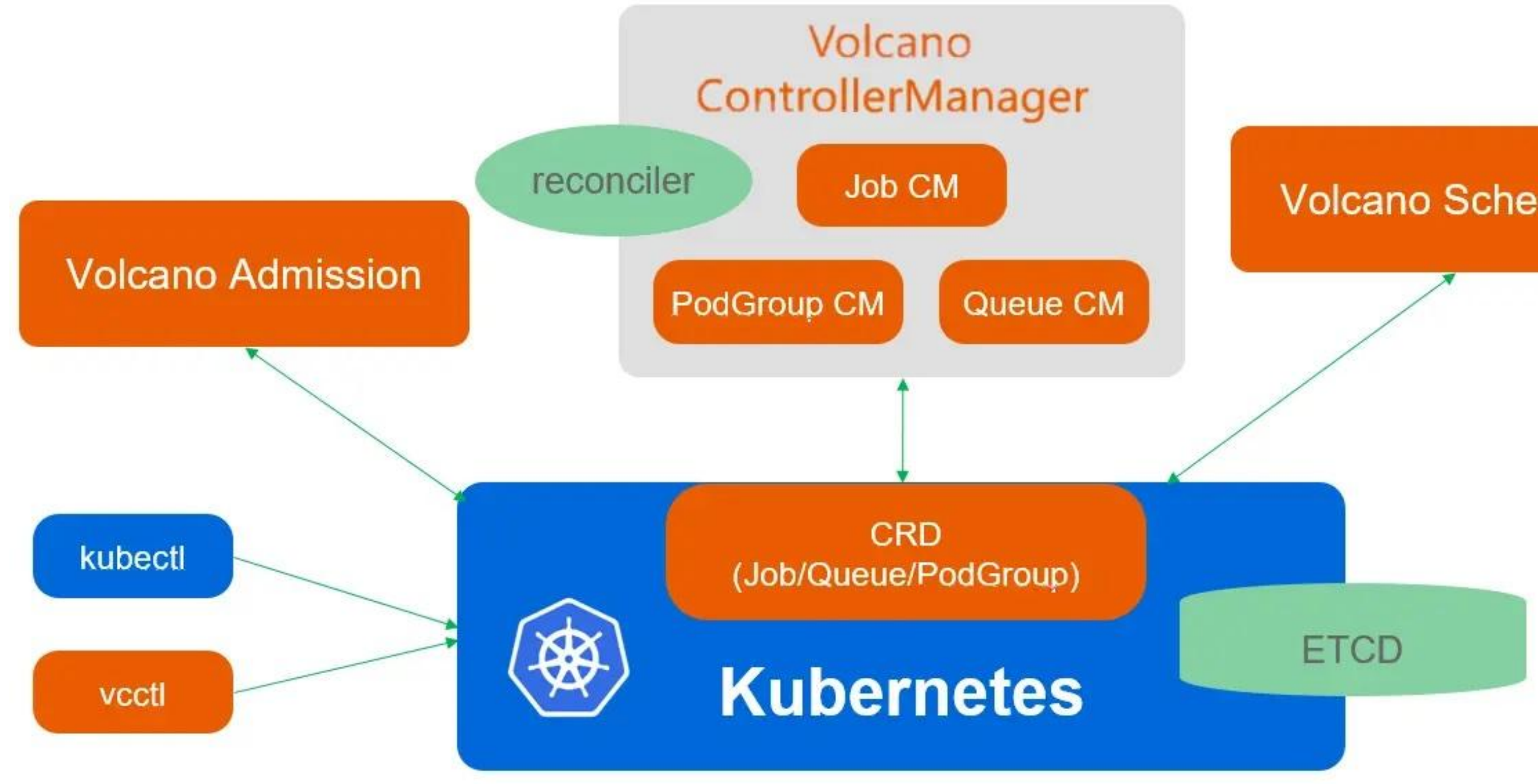
Why standard Kubernetes scheduling fails for ML workloads?


Problem A: The Fragmented Garage
Node 1
GPU
Job 1
8 GPUs
Job 3
Node 2
Job 2
Can't schedule despite 8 free GPUs.
Problem B: Wasted Cluster Capacity - Strict Quotas
Node 1
Quota: 4 GPUs
4 GPUs
Job 1
Job 3
4 slots are empty.
But the guard (Scheduler) says, "Sorry, your group has already used its 4 allocated spots," and doesn't let you in.
Problem C: Single-Cluster View
cluster 1: Scheduler 1
Job 1
Job 2
cluster x: Scheduler x
Job 1
Job 2
No Valet Between Lots!
6 GPUs
Job
Multi-Cluster awareness is absent as the scope is cluster-level.
Problem D: Handling Distributed Jobs
Node 1
Pod 1
Node 2
Pod 2
Running Job
Parking lot on hold: no single car moves in until the whole gang evacuates.
4 GPUs
Blocked Job

Custom Scheduling

Why Volcano?

1. Highly Extensible CNCF Design
   a. Native capabilities implemented as plugins
2. Good core framework
   a. Custom plugins can build a quota-based system driven by number of GPUs
3. OOTB support for PyTorch, TensorFlow workflows



Volcano Setup*

How do we build our scheduler?

1. Custom Plugin
   - Optimal Over Quota Behaviour
   - Project and Job Ordering for fairness
2. New Actions
   - Custom-Reclaim
   - Over-Quota Allocate
3. Behavior tuning for Gang, Messaging, etc.
4. New CRD : ClusterProject
   - Contains project quotas
   - Read by the scheduler

Cluster Project

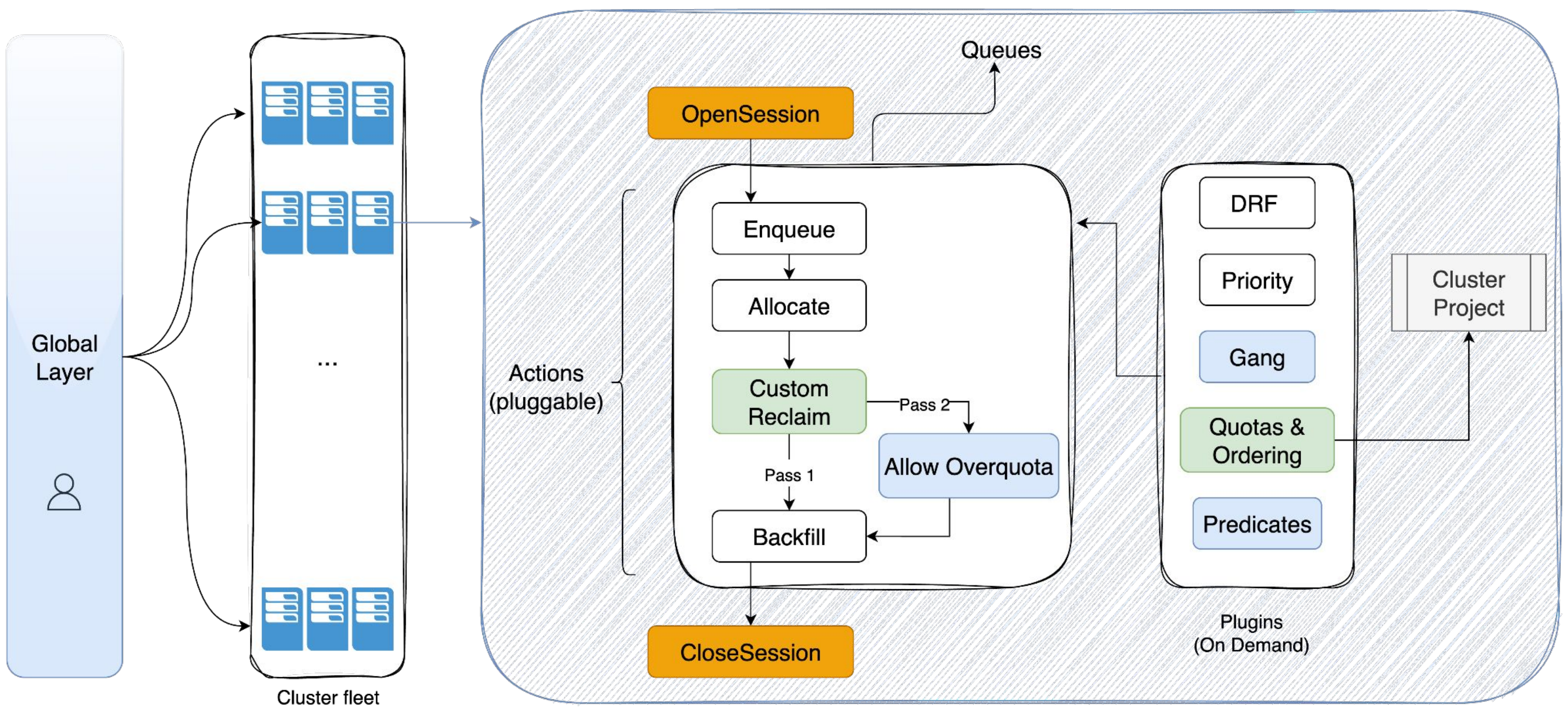
```
apiVersion: my.research.project/v1
kind: ClusterProject
spec:
  projectName: "ml-research-team"
  gpuQuota:
    "nvidia.com/gpu": 16
    "nvidia.com/a100": 8
  allowOverQuota: true
status:
  overQuotaGPUs: 4
  queuedJobs: 3
```

Proposed Solution

Two-Pass Scheduling Algorithm:

- **Pass 1 (Under-Quota):** Schedule jobs within project GPU limits using fair-share ordering
- **Pass 2 (Over-Quota):** Best-effort allocation of remaining resources to pending jobs

Topology Awareness:
All pods of a job are scheduled in the same node pool



Adobe

[*] https://www.infracloud.io/blogs/batch-scheduling-on-kubernetes/
https://www.cncf.io/blog/2021/02/26/volcano-collision-between-containers-and-batch-computing/